

# Uma Ferramenta para Geração de Esquemas Sintéticos de Bancos de Dados para Simulações P2P

Márcio de Carvalho Saraiva, Priscilla Kelly Machado Vieira, Carlos Eduardo Pires

Universidade Federal de Campina Grande (UFCG), Departamento de Sistemas e Computação (DSC), Campina Grande, PB, Brasil

{marcio,vieira,cesp}@dsc.ufcg.edu.br

**Abstract.** *To simulate P2P applications such as Peer Data Management Systems (PDMSs), simulators need to associate a database schema to each peer in the network. In general, the simulations involve a large number of data sources and find or create a large number of database schemas can be a tendentious and laborious task. This paper proposes a tool to automatically generate multiple synthetic database schemas for simulation purposes. The tool implements several algorithms to modify the structure of the synthetic database schemas.*

**Resumo.** *Para simular aplicações P2P como os Sistemas P2P de Gerenciamento de Dados, os simuladores necessitam associar um esquema de banco de dados a cada peer da rede. Em geral, as simulações envolvem um número elevado de fontes de dados, e encontrar ou criar uma grande quantidade de esquemas pode ser uma tarefa tendenciosa e cansativa. Este trabalho propõe uma ferramenta para geração automática de múltiplos esquemas de banco de dados sintéticos com propósito de simulação. A ferramenta implementa vários algoritmos para modificação da estrutura dos esquemas de bancos de dados sintéticos.*

## 1. Introdução

Simulação é a técnica mais comum para investigar aplicações *Peer-to-Peer* (P2P) tais como, Sistemas P2P de Gerenciamento de Dados (em inglês, *Peer Data Management Systems - PDMSs*) [Kantere *et al.*, 2008]. Para simular o processamento de consultas em PDMSs, os simuladores necessitam associar um esquema de banco de dados a cada *peer* da rede *overlay*. Este esquema representa os dados a serem compartilhados por um *peer* com os demais *peers* do sistema. Em geral, as simulações envolvem um número elevado de *peers*, de centenas a milhares. A tarefa de encontrar ou criar uma grande quantidade de esquemas de banco de dados pode ser tendenciosa e cansativa.

Este trabalho apresenta uma ferramenta para geração automática de esquemas de bancos de dados sintéticos com propósitos de simulação. A ferramenta funciona da seguinte forma: a partir de um esquema de banco de dados relacional suficientemente grande e complexo, enumerar múltiplos esquemas sintéticos e, em seguida, aplicar modificações semânticas na estrutura dos esquemas gerados. Os esquemas sintéticos são semanticamente similares entre si e, por conseguinte, úteis em simulações realísticas. Dependendo das modificações selecionadas, um nível maior ou menor de heterogeneidade pode ser introduzido nos esquemas sintéticos. O trabalho está estruturado da seguinte forma: a Seção 2 apresenta uma visão geral do processo de geração e modificação de esquemas sintéticos de bancos de dados, enquanto que a Seção 3 apresenta a ferramenta que implementa o processo descrito. Finalmente, a Seção 4 apresenta as conclusões e sugestões de trabalhos futuros.

## 2. Geração e Modificação de Esquemas Sintéticos de Bancos de Dados

A Figura 1 apresenta uma visão geral do processo de geração e modificação de esquemas sintéticos. Dado um esquema-base  $S$ , múltiplos esquemas sintéticos  $S_1, S_2, \dots, S_j$  são produzidos, onde cada esquema  $S_j$  corresponde a um subesquema modificado de  $S$ . Parâmetros podem ser usados para influenciar as características dos esquemas sintéticos a serem produzidos. O processo é não-determinístico visto que não há garantias de que dois conjuntos iguais de esquemas sintéticos serão produzidos após duas execuções do processo.

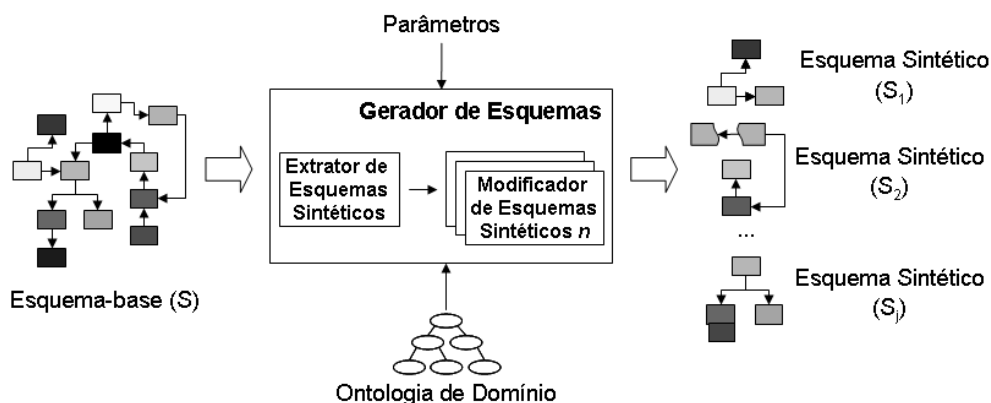


Figura 1. Visão geral do processo de geração e modificação de esquemas sintéticos.

O conjunto de operações para modificação de esquemas sintéticos inclui: (i) adição de colunas; (ii) remoção de colunas; (iii) substituição do nome de tabelas; (iv) substituição do tipo de dados de colunas; e (v) fusão de tabelas. Cada operação é implementada por um modificador de esquemas diferente. O processo é extensível no sentido de que novas operações podem ser acrescentadas. Cada modificador de esquemas recebe um esquema sintético  $S_j$  e modifica-o de acordo com sua função. Alguns modificadores de esquema como, por exemplo, (iii) utilizam uma ontologia de domínio [Guarino 1998] como recurso externo. As operações de modificação são aplicadas em uma ordem específica. Por exemplo, a substituição do nome das tabelas é executada após a adição e remoção de colunas. A ordem é determinada automaticamente, de acordo com os modificadores de esquemas selecionados pelo usuário.

## 3. Ferramenta para Geração e Modificação de Esquemas Sintéticos

A ferramenta<sup>1</sup> foi desenvolvida em Java. Para selecionar aleatoriamente os elementos (tabelas ou colunas) dos esquemas sintéticos a serem modificados, usamos a classe *Random* disponível no pacote *java.util*. Nesta versão, usamos a *Movie Ontology* [MO 2011] como ontologia de domínio. Conforme ilustrado na Figura 2, a ferramenta oferece uma interface gráfica que possibilita a configuração dos parâmetros de entrada. Um arquivo de *log* é criado a cada execução com sucesso. Este arquivo contém informações sobre o processamento gerado durante a execução da ferramenta tais como, tempo de execução e modificações realizadas nos esquemas.

Para utilização da ferramenta, o usuário deve conectar-se ao banco de dados que contém o esquema-base. A versão atual da ferramenta suporta conexões com o Sistema de Gerenciamento de Banco de Dados *Oracle 10g Express Edition*. Após a conexão com o banco de dados, o esquema-base é lido pela ferramenta e mostrado graficamente para o usuário. Para

<sup>1</sup> <http://code.google.com/p/schemagenerator2/>

este quesito, foi utilizada a tecnologia *JGraph* [JGraph 2011]. Em seguida, o usuário pode definir os parâmetros de entrada: número de esquemas sintéticos desejados, quantidade de tabelas em cada esquema e os modificadores de esquema a serem utilizados (Figura 2). Dependendo dos modificadores escolhidos, o usuário pode ainda definir o percentual de alterações por esquema e por tabela. Os esquemas sintéticos podem ser representados em dois formatos: relacional (SQL) ou ontologia (*Web Ontology Language* – OWL).

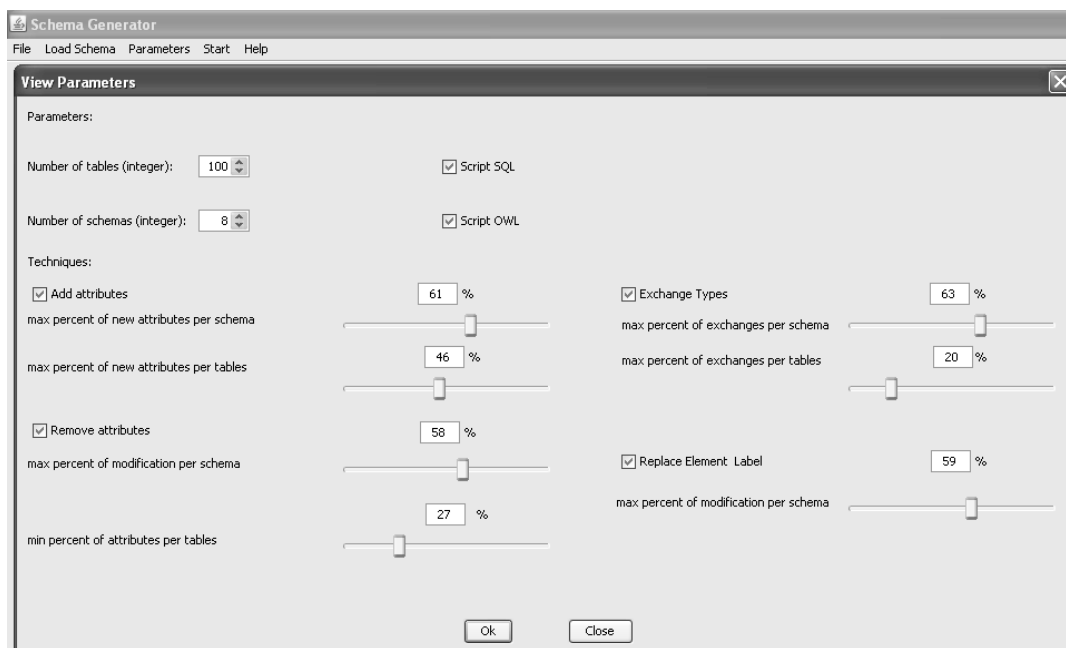


Figura 2. Tela de configuração de parâmetros da ferramenta.

#### 4. Conclusões e Trabalhos Futuros

Este trabalho propõe uma ferramenta para geração automática de múltiplos esquemas sintéticos de banco de dados. Os esquemas sintéticos produzidos pela ferramenta foram utilizados em um PDMS para simular o agrupamento semântico de *peers* de acordo com o esquema fornecido por cada *peer* [Pires 2009]. Como trabalhos futuros, sugerimos uma modificação no processo de geração de esquemas sintéticos no sentido de aceitar vários esquemas-base como entrada, ao invés de apenas um. Outra sugestão é a implementação de um algoritmo de modificação capaz de introduzir ruído nos esquemas sintéticos.

#### Bibliografia

- Guarino, N. (1998). Formal Ontology and Information Systems. In *1<sup>st</sup> International Conference on Formal Ontologies in Information Systems*, pages 3-15, Trento, Italy.
- JGraph. (2011). <http://www.jgraph.com/>, último acesso em 19/3/2011.
- Kantere, V., Tsoumakos, D., Sellis, T., Roussopoulos, N. (2008). GrouPeer: Dynamic Clustering of P2P databases. *Information Systems Journal*, Vol. 34, Issue 1, pages 62-86.
- OWL. (2011). <http://www.w3.org/TR/owl-features/>, último acesso em 19/3/2011.
- Pires, C. E. (2009). *Ontology-based Clustering in a Peer Data Management System*. Tese de Doutorado. Universidade Federal de Pernambuco (CIn/UFPE), Recife, Brasil.
- The Movie Ontology (MO). (2011). <http://www.movieontology.org/>, último acesso em 19/3/2011.